

IN THE CLAIMS

Please cancel claims 10-11 and 15-16.

Please amend the claims as follows.

- 1 1. (Currently Amended) An apparatus comprising:
 - 2 at least one processor;
 - 3 a memory coupled to the at least one processor;
 - 4 a database table residing in the memory; [[and]]
 - 5 a cardinality estimator residing in the memory and executed by the at least one
 - 6 processor, the cardinality estimator estimating cardinality of an intermediate dataset that
 - 7 satisfies a query to the database table in a manner that accounts for data skew in the
 - 8 database table; and
 - 9 a query optimizer residing in the memory that uses the estimated cardinality from
 - 10 the cardinality estimator to optimize the query.

- 1 2. (Currently Amended) The apparatus of claim 1 further comprising a frequent values
2 list residing in the memory that contains a list of values in the database table, each value
3 having a corresponding frequency, wherein the cardinality estimator estimates the
4 cardinality of the intermediate dataset by determining whether a frequency corresponding
5 to a value exceeds a predetermined threshold, and if the frequency exceeds the
6 predetermined threshold, accounting for the corresponding value, and if the frequency
7 does not exceed the predetermined threshold, using a formula to estimate the cardinality
8 of the intermediate dataset, the formula accounting for data skew in the database table by
9 subtracting the frequency of all values above the predetermined threshold in the frequent
10 ~~value table~~ values list that satisfy the query from the total number of ~~columns~~ rows in the
11 database table.

1 3. (Original) The apparatus of claim 2 wherein the cardinality estimator estimates the
2 cardinality Ca' of the intermediate dataset using the formula:

3

$$Ca' = P + M \left(1 - \left(1 - \frac{1}{M} \right)^Y \right)$$

4 where

5

$$M = Ca - (P+Q)$$

6 P = number of distinct values in the frequent values list above the
7 predetermined threshold that satisfy the query;

8 Q = number of distinct values in the frequent values list above the
9 predetermined threshold that do not satisfy the query;

10 Ca = cardinality of the database table;

11 $Y = X - Fi$;

12 X = number of rows in the intermediate dataset; and

13 Fi = sum of frequencies of values in the frequent values list above the
14 predetermined threshold that satisfy the query.

1 4. (Currently Amended) An apparatus comprising:
2 at least one processor;
3 a memory coupled to the at least one processor;
4 a database table residing in the memory;
5 a frequent values list residing in the memory that contains a list of values in the
6 database table, each value having a corresponding frequency; and
7 a cardinality estimator residing in the memory and executed by the at least one
8 processor, the cardinality estimator estimating cardinality of [[the]] an intermediate
9 dataset for a query to the database table using the following formula:

10
$$Ca' = P + M(1 - (1 - \frac{1}{M})^Y)$$

11 where

12 $M = Ca - (P+Q)$

13 P = number of distinct values in the frequent values list above [[the]] a
14 predetermined threshold that satisfy the query;

15 Q = number of distinct values in the frequent values list above the
16 predetermined threshold that do not satisfy the query;

17 Ca = cardinality of the database table;

18 $Y = X - Fi$;

19 X = number of rows in the intermediate dataset; and

20 Fi = sum of frequencies of values in the frequent values list above the
21 predetermined threshold that satisfy the query; and

22 a query optimizer residing in the memory that uses the estimated cardinality of the
23 intermediate dataset to optimize the query.

1 5. (Currently Amended) A method for optimizing a query to a database table by
2 estimating cardinality of an intermediate dataset that results from processing [[a]] the
3 ~~database query on a database table~~, the method comprising the steps of:

- 4 (A) evaluating the query; [[and]]
5 (B) estimating cardinality of the intermediate dataset using a formula that
6 accounts for data skew in the database table; and
7 (C) using the cardinality estimate in step (B) to optimize the query.

1 6. (Currently Amended) The method of claim 5 wherein step (B) includes the steps of:

- 2 selecting a value in a frequent values list that contains a list of values in the
3 database table, each value having a corresponding frequency;
4 if the selected value has a corresponding frequency that exceeds a predetermined
5 threshold, incrementing the cardinality estimate by one; and
6 if the frequency does not exceed the predetermined threshold, using a formula to
7 estimate the cardinality of the intermediate dataset, the formula accounting for data skew
8 in the database table by subtracting the frequency of all values above the predetermined
9 threshold in the frequent value table that satisfy the query from the total number of
10 ~~columns~~ rows in the database table.

1 7. (Currently Amended) The method of claim 6 wherein ~~the cardinality estimator~~
2 ~~estimates the step of estimating~~ the cardinality Ca' of the intermediate dataset in step (B)
3 ~~using~~ uses the formula:

4

$$Ca' = P + M \left(1 - \left(1 - \frac{1}{M}\right)^Y\right)$$

5 where

6 $M = Ca - (P+Q)$

7 P = number of distinct values in the frequent values list above the
8 predetermined threshold that satisfy the query;

9 Q = number of distinct values in the frequent values list above the
10 predetermined threshold that do not satisfy the query;

11 Ca = cardinality of the database table;

12 $Y = X - Fi$;

13 X = number of rows in the intermediate dataset; and

14 Fi = sum of frequencies of values in the frequent values list above the
15 predetermined threshold that satisfy the query.

- 1 8. (Currently Amended) A method for optimizing a query to a database table by
2 estimating cardinality of an intermediate dataset that results from processing [[a]] the
3 database query on a database table, the method comprising the steps of:
4 (A) evaluating the query; [[and]]
5 (B) estimating the cardinality Ca' of the intermediate dataset using the formula:

6

$$Ca' = P + M \left(1 - \left(1 - \frac{1}{M}\right)^Y\right)$$

7 where

8 $M = Ca - (P+Q)$

9 P = number of distinct values in the frequent values list above [[the]] a
10 predetermined threshold that satisfy the query;

11 Q = number of distinct values in the frequent values list above the
12 predetermined threshold that do not satisfy the query;

13 Ca = cardinality of the database table;

14 $Y = X - Fi$;

15 X = number of rows in the intermediate dataset; and

16 Fi = sum of frequencies of values in the frequent values list above the
17 predetermined threshold that satisfy the query; and

18 (C) using the cardinality estimate in step (B) to optimize the query.

- 1 9. (Currently Amended) A program product comprising:
- 2 (A) cardinality estimator estimating cardinality of an intermediate dataset that
- 3 satisfies a query to a database table in a manner that accounts for data skew in the
- 4 database table;
- 5 (B) a query optimizer that uses the estimated cardinality from the cardinality
- 6 estimator to optimize the query; and
- 7 (B) computer-readable signal bearing (C) recordable media bearing the cardinality
- 8 estimator and the query optimizer.
- 1 10. (Cancelled)
- 1 11. (Cancelled)
- 1 12. (Currently Amended) The program product of claim 9 wherein the cardinality
- 2 estimator evaluates a frequent values list that contains a list of values in the database
- 3 table, each value having a corresponding frequency, wherein the cardinality estimator
- 4 estimates the cardinality of the intermediate dataset by determining whether a frequency
- 5 corresponding to a value exceeds a predetermined threshold, and if the frequency exceeds
- 6 the predetermined threshold, accounting for the corresponding value, and if the frequency
- 7 does not exceed the predetermined threshold, using a formula to estimate the cardinality
- 8 of the intermediate dataset, the formula accounting for data skew in the database table by
- 9 subtracting the frequency of all values above the predetermined threshold in the frequent
- 10 value table values list that satisfy the query from the total number of columns rows in the
- 11 database table.

1 13. (Currently Amended) The program product of claim 12 wherein the cardinality
2 estimator estimates the cardinality Ca' of the intermediate dataset using the formula:

3

$$Ca' = P + M \left(1 - \left(1 - \frac{1}{M} \right)^Y \right)$$

4 where

5

$$M = Ca - (P+Q)$$

6 P = number of distinct values in the frequent values list above the
7 predetermined threshold that satisfy the query;

8 Q = number of distinct values in the frequent values list above the
9 predetermined threshold that do not satisfy the query;

10 Ca = cardinality of the database table;

11 $Y = X - Fi$;

12 X = number of rows in the intermediate dataset; and

13 Fi = sum of frequencies of values in the frequent values list above the
14 predetermined threshold that satisfy the query.

1 14. (Currently Amended) A program product comprising:

2 (A) a cardinality estimator that estimates cardinality of [[the]] an intermediate
3 dataset for a query to a database table using the following formula:

4

$$Ca' = P + M \left(1 - \left(1 - \frac{1}{M} \right)^Y \right)$$

5 where

6 $M = Ca - (P+Q)$

7 P = number of distinct values in [[the]] a frequent values list above [[the]]
8 a predetermined threshold that satisfy the query;

9 Q = number of distinct values in the frequent values list above the
10 predetermined threshold that do not satisfy the query;

11 Ca = cardinality of the database table;

12 Y = X - Fi;

13 X = number of rows in the intermediate dataset; and

14 Fi = sum of frequencies of values in the frequent values list above the
15 predetermined threshold that satisfy the query;

16 (B) a query optimizer that uses the estimated cardinality from the cardinality
17 estimator to optimize the query; and

18 (B) computer-readable signal bearing (C) recordable media bearing the cardinality
19 estimator and the query optimizer.

1 15. (Cancelled)

1 16. (Cancelled)